

# **Digital Projects Guidelines**

**Arizona State Library, Archives and Public Records**

Phoenix, Arizona  
January 2006

Version 3.0

## **Table of Contents**

<b>INTRODUCTION .....</b>	<b>2</b>
<b>ABOUT THE ARIZONA MEMORY PROJECT .....</b>	<b>2</b>
<b>ORIENTING YOUR STAFF.....</b>	<b>5</b>
<b>PLANNING YOUR PROJECT .....</b>	<b>5</b>
<b>EQUIPMENT AND SOFTWARE.....</b>	<b>8</b>
<b>DIGITIZING YOUR MATERIALS.....</b>	<b>10</b>
<b>DIGITIZATION GUIDELINES.....</b>	<b>12</b>
<b>DESCRIBING AND CATALOGING YOUR MATERIALS.....</b>	<b>13</b>
<b>USING A DATABASE.....</b>	<b>15</b>
<b>PRESENTING YOUR MATERIALS .....</b>	<b>17</b>
<b>LONG-TERM STORAGE AND MAINTENANCE OF THE DIGITAL FILE .....</b>	<b>19</b>
<b>APPENDIX A: FREQUENTLY ASKED QUESTIONS .....</b>	<b>21</b>
<b>APPENDIX B: RECOMMENDED RESOURCES .....</b>	<b>23</b>

## **INTRODUCTION**

The Arizona State Library, Archives and Public Records (ASLAPR) provides coordination, assistance and support for Arizona's historical and cultural institutions as they engage in digitization projects. These projects, when successful, offer users exciting new levels of access to the unique and priceless collections in our institutions. The guidelines have been developed to assist you in planning and completing digitization projects, and to provide our patrons the broadest possible public access to all cultural collections in Arizona.

Adhering to these guidelines will enable your institution to provide the best quality of service. It also creates a foundation for the exciting potential to inter-connect many collections electronically in statewide, regional or national consortia.

This document recommends minimum guidelines, developed primarily to support public access. It targets individuals considering their first project or who are still relative novices in scanning and digitization efforts. Institutions and individuals with significant digitization experience may find these guidelines rudimentary, and are encouraged to build and add detail as appropriate. Institutions with higher budget and staffing levels may choose to exceed these recommendations and provide higher quality scans and more detailed item descriptions. Though the minimum standards are relatively stable, these guidelines are dynamic and will be continually refined and updated using feedback from successful local, regional, national and international projects. These guidelines also recognize the problematic nature of long-term preservation of electronic records. The preservation of items in their original forms is a key aspect of any digital project.

## **ABOUT THE ARIZONA MEMORY PROJECT**

If you want your patrons and staff to view, manage and access your digital resources online, consider the database software used by the Arizona Memory Project. The software and storage is available at no cost to libraries, archives, and museums in Arizona, courtesy of the Arizona State Library, Archives and Public Records.

Established by the Arizona State Library, Archives and Public Records, the Arizona Memory Project is a digital library initiative that is specifically designed to store, describe, display and administer digital collections from archives, libraries, museums, historical societies and other Arizona cultural institutions. The aim of this initiative is to create a central online repository to provide access to unique digital content representative of the history and culture of Arizona. The project can handle a multitude of collections consisting of many different types of file formats for items including photographs, newspapers, maps, manuscripts, oral histories and more. Patrons can view your collection from any computer with Internet access.

The project adheres to the Arizona State Library, Archives and Public Records Metadata Guidelines, which is based on Dublin Core metadata, and uses a software package that provides a simple and efficient way to get your digital resources online.

The Arizona Memory Project is supported with funds granted by the Arizona State Library, Archives and Public Records Agency under the Library Services and Technology Act, which is administered by the Institute of Museum and Library Science.

To learn more about the Arizona Memory Project, please visit <http://azmemory.lib.az.us>. To learn more about the software, visit <http://www.contentdm.com> or contact the Arizona Memory Project Coordinator at [azmemory@lib.az.us](mailto:azmemory@lib.az.us). To learn more about joining Arizona Memory Project, contact Library Development Division at [ldd@lib.az.us](mailto:ldd@lib.az.us).

## ARIZONA STATE LIBRARY, ARCHIVES AND PUBLIC RECORDS STAFF CONTACTS AVAILABLE FOR ASSISTANCE

<b>Computers/Internet</b>		
Technology services	(602) 542-4035	helpdesk@lib.az.us
<b>Photographic Applications</b>		
<b>Digitizing/Scanning Techniques</b>		
Laurie Devine	(602) 542-4159	ldevine@lib.az.us
<b>Library Applications</b>		
<b>Description/Indexing/Cataloging/Metadata/Crosswalks</b>		
Mary Alice Cicerale	(602) 542-4046	macicerale@lib.az.us
Sara Muth	(602) 542-4046	smuth@lib.az.us
<b>Museum Applications/Cataloging</b>		
Brenda McLain	(602) 542-4675	bmclain@lib.az.us
<b>Archival Applications</b>		
Wendi Goen	(602) 542-4159	wgoen@lib.az.us
<b>Database Design and Implementation</b>		
Laurie Devine	(602) 542-4159	ldevine@lib.az.us
<b>CONTENTdm / Arizona Memory Project</b>		
Marisa Ramirez	(602) 542-3701	mramirez@lib.az.us
Richard Prouty	(602) 258-5053	rprouty@lib.az.us
<b>Usability/Developing Templates/ Websites</b>		
Richard Pearce-Moses	(602) 364-0265	rpm@lib.az.us
<b>File Formats/Long Term Storage of Electronic Records</b>		
Bill James	(602) 542-3741	bjames@lib.az.us
<b>Preservation of Original Sources</b>		
Michael McColgin (documents)	(602) 542-4159	mimccol@lib.az.us
Laurie Devine (photographs)	(602) 542-4159	ldevine@lib.az.us
Brenda McLain (artifacts)	(602) 542-4675	bmclain@lib.az.us
<b>Paper Copies From Selected Web Sites If You Do Not Have Internet Access</b>		
Wendi Goen	(602) 542-4159	wgoen@lib.az.us

# GETTING STARTED

## ORIENTING YOUR STAFF

Educating your staff and administrators about the issues involved in successfully planning, implementing, and maintaining a digital collection is extremely important. You are encouraged to draw upon the expertise of our Agency staff as you develop and implement your programs. The table on the previous page provides information on Arizona State Library, Archives and Public Records staff available to assist you in understanding these guidelines.

Valuable current information about digitizing is also available on the Web to provide additional background and examples of scope, organization, design, and cataloging detail. As you become involved in your projects, we strongly recommend consulting the information sources listed in APPENDIX B: RECOMMENDED RESOURCES.

## PLANNING YOUR PROJECT

In planning your project, you should first determine why you wish to digitize your collection. The best reason to digitize involves improving access to your unique collections. It increases the visibility of your collection and provides an attractive alternative to handling of the originals.

It is critical that you establish guidelines for the scope and content of both your immediate work and a long range plan. You should begin by asking yourself questions about why you are undertaking your project.

### ELEMENTS TO CONSIDER IN PLANNING YOUR PROJECT

**Mission and Scope of the Project:** You should consider the purpose and audience of the project and the information to be conveyed to that audience. Developing written policies that address these issues will help focus the project in your mind and help ensure that everyone involved has a common understanding of the project. The policies can be used as a "reality check" by soliciting input from project staff, administrators, users, and others.

It is not essential to spend a great deal of time on these policies, especially when an organization is developing its first digital project or Website, or a new collection of pages for an existing site. These policies should be modified as the project evolves and as you acquire more experience with how the project can meet your organization's needs. Policies on mission and scope of the project should consider the following:

**What Are Your Goals?** What do you want the project to accomplish? Are you expanding outreach, improving access to a wider audience that can't easily come to your site or can't come during your regular hours? Do you want to improve preservation of the materials by reducing handling of originals? How do your project goals fit into your organization's larger strategic plan?

**Who Will Use The Information?** It is important to design for your intended audience and keep their needs in mind as you plan and develop your project. Will they be educators, students, researchers, family historians or the general public? Do not forget the purpose of your project is to communicate with your audience.

## EFFECTIVE PROJECT PLANS ADDRESS THE FOLLOWING QUESTIONS

**How Will They Use the Information?** What kind of information will the users want, and how will they want it packaged? How much do they know about the organization of library collections, museums, or archives? Will they phrase general or specific research questions? What are their computer and Internet skills? Do they need a lot of explanation and help, are they sophisticated users, or both? What are the users' computer resources?

**What Resources Are Available?** Acquiring the hardware and software for the system is only one portion of the process. Preservation of original materials after digitizing, managing back-up files, description and cataloging take time and effort, as do the design and development of your project. Time estimates indicate that two hours of technology staff or maintenance may be required for every hour devoted to the preparation of an online digital library. Some estimates range as high as 20 to 1! In creating your original plan, offer a realistic estimate of the time required per image file to house, accession, scan, resize, create a catalog record and apply the file to your system. Plan for long term maintenance and migration of your digital files. Don't assume existing resources devoted to an off-line equivalent will be able to absorb the additional work to support the Web service. On-line access to collections can significantly increase requests for research assistance, reproductions of photographs and other services.

**Who Will Do the Work?** Photographic experience is very helpful in learning to scan and adjust image files. Likewise, librarians, archivists, and curators have experience helpful in composing cataloging records (called metadata). Technical support and screen design are other specialized skills helpful for successful projects. If more than one person works on the project, it is essential that staff work closely together from the initial planning. Also, cataloging and digitizing are labor intensive and may require a skill level above that of minimum wage employees. Plan for training if current staff does not have the necessary skills.

**Who Is Responsible for the Project?** As noted above, any large project will have many people working together. Ultimately, one individual must be designated to manage the project and make final decisions. This individual should have a strong grasp of the mission and scope of the entire project and an understanding of technological aspects, information organization, and interface design.

**How Will Users Gain Access?** Will your system be for in-house use by patrons and staff or will you make material available online? Issues such as cultural sensitivity and copyright become especially important when image files are posted online. In-house use generally considered "Fair Use" within your institution may not be "Fair Use" outside of your institution on the Internet. Design questions include: How will you protect your files from being downloaded off the Internet, or do you care? What software will you use? Who will create and maintain the in-

house system and/or web site? Decisions about file size will be influenced by how your users will get access to the information. For example, small file sizes take less time to load and are preferred if your users use modems.

**Copyright Issues, Who Owns the Rights?** Generally, the law provides that a published item is protected under copyright for at least 75 years, unless the materials are in the public domain. In general, United States government publications are in the public domain; however, copyright may apply to state and local works. Inquire with the appropriate government agency about possible copyright protection for its materials. Unless rights are transferred, copyright remains with the producer of the item, his/her descendants, or designee, such as a publisher or employer, and must be transferred in writing. Unpublished material and that published after January 1, 1923, is protected under current copyright law and you should only reproduce such material with permission of the copyright holder. Donor agreements should include transfer of rights for unpublished material, and published material produced after January 1, 1923. If you are not certain about your institution's rights to publish an item, you should verify ownership and permissions before doing so. The U.S. Copyright Office is examining issues raised by "orphan works," i.e., copyrighted works whose owners are difficult or even impossible to locate. Concerns have been raised that the uncertainty surrounding ownership of such works might needlessly discourage subsequent creators and users from incorporating such works in new creative efforts, or from making such works available to the public. Refer to the U.S. Copyright Office for more information on "orphan works".

**What Material Will You Digitize?** The size and physical conditions of the items will determine what equipment and software will be necessary for your project. Photographs and fragile documents require significant staff attention and possibly specialized software, such as Photoshop. Grouping material can save time later in the project. Decide how you plan to digitize your collections: the most popular collections first, everything in a collection, regardless of value of specific items, or a representative sampling from many collections.

**Preservation:** After scanning it is important to correctly house and store your original, paper or source material, according to standard preservation practices. A good digitization program provides an excellent opportunity to preserve and maintain the original as a part of its plan. Photographs, documents, and artifacts can be properly rehoused, accessioned and stored in an improved environment. Generally this adds little to the cost of the project, and will add to the life of the collection. Preservation of original material is required in all projects supported by our Agency. If your materials are not already properly housed, it could add significant time and expense to the project.

**Will You Purchase or Lease Hardware and Software?** If purchasing, be prepared for future expenses to upgrade the system. If leasing, ensure you have an exit strategy.

**What Physical Facilities Are Available?** Scanning projects require adequate space for computers, scanning devices, printers, peripherals, people and preservation, even if there is only one person working on the project. Plan for an appropriate amount of space in an environmentally appropriate location.



**How Will You Store and Maintain the Information?** The most popular current options are tape, compact disc and on a fixed disc. If you choose to store your image or text files on a hard drive, store the files in formats recommended on page 12. Plan for adequate back-up of data files, and for sufficient storage space for file servers if material will be available on the Internet.

## **EQUIPMENT AND SOFTWARE**

Effectively selecting computer hardware and software is essential for a successful project. These guidelines offer general specifications for hardware (computer equipment and scanners), software (computer programs for digitizing and organizing files into a database) and storage media for your files. Additional consultation support is available from our Agency staff.

The goals of your digitization project will determine what hardware and software you need to successfully complete the project. Scanning, image manipulation, attaching text information, indexing, image display and long-term storage can require different equipment and user expertise. What materials are to be scanned and how those images are to be used will impact equipment choices. These selection guidelines are intended as general recommendations. Digitization is an area that is evolving and we strongly recommend that you research the current literature in print and online for the latest developments before purchasing any equipment.

### **Before purchasing equipment consider:**

What can your staff handle?

What can your current local technology support?

Are you scanning photo and/or art images or documents?

How will price impact your selection? (affordability vs. performance)

Think about your users and potential users, how will they use the images?

Do you plan to store the images in a common (standardized) file format?

Remember that digitized image files are potentially very large, do you have enough storage capacity?

### **Minimum specifications for hardware selection**

PC/Macintosh

Highest processing speed you can afford

512MB RAM

40 GB Hard Drive or larger

17" – 21" Monitor

CD-ROM drive

CD-RW and/or DVD writer

Color/BW printer, preferably laser or capable of using archival inks

Scanner

## **Scanners**

Flatbed scanners are recommended for most digitization projects. Flatbed scanners can be used for photographs in sizes up to 12x18 (depending upon the size of the bed). Scanners may be able to support transmitted light (for transparencies or negatives) depending on current and projected future collection holdings. A slide scanner may be a desired addition if your collection includes large numbers of slides or 35mm/small format negatives

Flatbed scanner                      (minimum resolution 1200 dpi with 32 bit depth)

Optional Slide scanner              (minimum resolution 2400 dpi with 32 bit depth)

The digital camera promises to expand the digitization tool kit. Low resolution, relatively inexpensive digital cameras may be used to produce thumbnail images for publication on the Web. High-resolution digital cameras are becoming a viable option for digitization projects, but rapid changes in technology make it difficult to provide specific digital camera hardware specifications. Consult the web or contact Arizona State Library, Archives and Public Records staff for more information. Special cases may require the use of a digital camera on copy stand to permit digitizing flat art and 3-D objects and artifacts.

The Records Management Division of this Agency offers vendor service to Arizona institutions for textual document imaging (not photographs). Contact Bill James (phone #602.542.3741; email: [bjames@lib.az.us](mailto:bjames@lib.az.us)) for information. Other private vendors and institutions in Arizona and the Western U.S. also offer this service.

## **General specifications for software selection**

Scanning software comes with the scanner itself, and is critical. A quality scanner without good software can be ineffective and vice-versa. Engage in serious comparison before acquiring a scanner. Image editing software (Adobe PhotoShop full version preferred over Adobe PhotoShop LE) is a necessity and must be purchased separately. Ensure that scanning software is “descreen” capable. (Descreening refers to the process of removing a moiré pattern usually associated with scanning an image that has been through the four color CMYK process.) This software capability allows for the scanning of images which have been printed in books or magazines.

## **General specifications for storage media selection**

Digital files are susceptible to data corruption resulting from software malfunctions and hardware failures. To combat the loss of electronic information, use at least two different storage media, verify file integrity immediately after creating offline copies and then check periodically for signs of corruption. The most common methods of file storage are described below.

CD-Writable (CDR) is recommended for storage of data files. As DVD burners become more common, the use of DVD discs may be another storage option for you. Digital files should be

regularly checked for quality control and file integrity, and should be migrated to new media as appropriate. CD-Rewritables (CDRW) are acceptable for working files, but are inappropriate for long term storage.

External hard drives offer large storage capacity and can be used to back up existing files. Be aware that external hard drives are as susceptible to mechanical failure as the hard drive inside your computer, and because the external hard drive must be plugged into your computer for backups, it is vulnerable to any catastrophe that destroys your computer.

CD-R for capacity up to 700MB

DVD-R /DVD+R for capacity from 4.7 GB to 8.54 GB

External hard drives start at 40GB and go up to a full terabyte (1,024 GB)

## **DIGITIZING YOUR MATERIALS**

Providing electronic visual access requires electronic (digital) surrogates of the photographs, documents and records that can be made available to patrons on a dedicated computer, CD-ROM or on the Internet/Web. The following guidelines provide minimum benchmarks for scanning, file format, quality of resolution of the scanned image, description and cataloging, database organization and screen display.

These are the recommended *minimum* guidelines for Arizona institutions involved in digitization projects in cooperation with the Arizona State Library, Archives and Public Records. Image files should be saved and scanned in at least three file sizes. Resizing images or interpolating to a higher resolution from a lower resolution for the purposes of transfer is not acceptable. Image files should not be cropped or altered, except for the removal of extraneous borders.

Proper image calibration based upon grayscale or color measurements (not screen display) is presumed and is a critical component of the scanning process. One approach would be to use a commercially available target for image calibration. Targets provide a mechanism for benchmarking the capture process. In simple terms, a target is a sample with known characteristics that can be used to establish a baseline or standard by which to assess capture. Your project should adopt a policy on what targets and scales to use, when to use them, and how their quality will be controlled. Two types of targets are commonly used: edge and detail (resolution) targets and color or grayscale charts. After targets are scanned, they are evaluated with a software program.

These guidelines are for materials smaller than 11 x 14".

*Master files* are the largest files maintained and should not be altered, compressed or resized. They provide the source files for all other files and should remain in a controlled environment, with limited use. They also provide files for high-quality publications, but generally not for everyday staff use.

*Access files* are generally the largest day-to-day use. They may be “uploaded” on web pages, used for patron requests and local printing. They are good quality, but still of manageable size.

*Thumbnail files* are very small. They can be collected into databases or used for web page display. They load quickly, but are very small and unsuitable for printing.

Institutions requiring uniform image (not file) sizes should use the pixel on the long dimension option. DPI (dots per inch) scanning will produce varied image sizes consistent with the original.

Photographic image manipulation software, such as Photoshop, Photosuite, Paintshop Pro each have different ways of calculating the level of JPEG compression. The following recommendations are based upon a 1 (lowest) to 10 (highest) quality system which is most common. Other software should consider the “low,” “medium,” and “high” quality indicators on the JPEG window.

Thumbnails of maps, drawings and bitonal images should only be produced when they provide visual information while viewing. For example, if the outline of the map is useful in the small size, that would justify creating a thumbnail. However, if the map loses detail, a thumbnail may be unusable and irrelevant,

Textual files which contain artwork, graphs, and other non-textual matter should be treated as “Documents” and those standards should apply. A textual file should contain print & numeric characters.

# DIGITIZATION GUIDELINES

## Arizona State Library, Archives and Public Records

### Minimum Scanning Guidelines

These are the *minimum* guidelines for Arizona institutions involved in digitization projects in cooperation with the Arizona State Library, Archives and Public Records. Image files should be saved and scanned in at least three file sizes. Image files should not be cropped or altered, except for the removal of extraneous borders. Proper image calibration based upon grayscale measurements (not screen display) is presumed and is a critical component of the scanning process. These guidelines are for materials not to exceed 11 x 14".

<b>File type</b>	<b>Textual</b>	<b>Photographic</b>	<b>Documents</b>	<b>Maps, Drawings, Bitonal</b>
<b>Master files</b>	<b>Scan resolution</b> 200-300 dpi grayscale TEXT only min. 11" on long dimension  <b>File format &amp; resolution</b> Uncompressed TIFF Intel (IBM) byte order 200 dpi at original size	<b>Scan resolution</b> 4000 pixels on long dimension OR 600 dpi  <b>File format &amp; resolution</b> Uncompressed TIFF Intel (IBM) byte order 4000 pixels on long dimension OR 600 dpi at original size	<b>Scan resolution</b> 4000 pixels on long dimension OR 600 dpi  <b>File format &amp; resolution</b> Uncompressed TIFF Intel (IBM) byte order 4000 pixels on long dimension OR 600 dpi at original size	<b>Scan resolution</b> 300 dpi  <b>File format &amp; resolution</b> Uncompressed TIFF Intel (IBM) byte order 200-300 dpi at original size
<b>Access files</b>	<b>File format</b> 8 bit grayscale JPEG 4-6 on a 1/10 scale (medium)  <b>File resolution</b> 200 dpi unaltered image size	<b>File format</b> 8 bit grayscale, 24 bit color JPEG 8-10 on a 1/10 scale (high)  <b>File resolution</b> 300 dpi unaltered image size	<b>File format</b> 8 bit grayscale, 24 bit color JPEG 8-10 on a 1/10 scale (high)  <b>File resolution</b> 300 dpi unaltered image size	<b>File format</b> 8 bit grayscale, 24 bit color JPEG 8-10 on a 1/10 scale (high)  <b>File resolution</b> 200-300 dpi unaltered image size OR reduced to equivalent of 8x10"
<b>Thumbnails</b>	<b>Not recommended for textual files</b>	<b>File format &amp; resolution</b>  4 bit grayscale, 8 bit color JPEG 4-5 on a 1/10 scale (medium) 72 dpi or 94 dpi	<b>File format &amp; resolution</b>  4 bit grayscale, 8 bit color JPEG 4-6 on a 1/10 scale (medium) 72 dpi or 94 dpi	<b>Optional for Bitonal, maps &amp; drawings</b>  4 bit grayscale, 8 bit color JPEG 4-6 on a 1/10 scale (medium) 72 dpi or 94 dpi

#### Notes:

*Master files* are the largest files maintained and should not be altered, compressed or resized. They provide the source files for all other files and should remain in a controlled environment, with limited use. They also provide files for high-quality publications, but generally not for everyday staff use.

*Access files* are generally the largest day-to-day use. They may be "uploaded" on web pages, used for patron requests and local printing. They are good quality, but still of manageable size. Access files for use on the web may be scanned at 640w x 480h pixels.

*Thumbnail files* are very small. They can be collected into databases or used for web page display. They load quickly, but are very small and unsuitable for printing.

Institutions requiring uniform image (not file) sizes should use the pixel on the long dimension option. DPI scanning will produce varied image sizes consistent with the original. Thumbnails of maps, drawings and bitonal images should only be produced when they provide visual information while viewing. For example, if the outline of the map is useful in the small size, that would justify creating a thumbnail. Thumbnails of maps, drawings and bitonal images should only be produced when they provide visual information while viewing. For example, if the outline of the map is useful in the small size, that would justify creating a thumbnail.

However, if the map loses detail, a thumbnail may be unusable and irrelevant.

Textual files which contain artwork, graphs, and other non-textual matter should be treated as "Documents" and those standards should apply. A textual file should contain print & numeric characters.

Photographic image manipulation software, such as Photoshop, Photosuite, Paintshop Pro each have different ways of calculating the level of JPEG compression. We've offered the recommendations based upon a 1 (lowest) to 10 (highest) quality system which is most common. Other software should consider the "low," "medium," and "high" quality indicators on the JPEG window.

## Organizing the Digital Files

Physical organization of your digital files is very important. You should develop a structure to guide naming and organizing files during the planning process.

Files can be named and organized in a hierarchy. Similar types of files should be grouped into subdirectories. Related subdirectories can then be grouped into directories. Typically, computer files names are organized from directory to subdirectories to file with the names separated by a slash"/". Here is an example of a file name with directory and subdirectory structure:

/photograph/1999/buildings/1234pierceave/front.gif

Directory	/photographs
Subdirectory (year)	/1999
Subdirectory (category)	/buildings
Subdirectory (title)	/1234pierceave
Files	/front.gif

The strategies for naming and organizing files will vary across institutions, but the name of each directory and file should be meaningful. It is very important that the processes for naming and organizing are clearly documented to make it possible for new files to be added by new staff or volunteers as the project is maintained and grows over time.

## DESCRIBING AND CATALOGING YOUR MATERIALS

The importance of consistent, comprehensive description and cataloging is frequently one of the most overlooked aspects of the project. Digital materials need to be described and cataloged effectively to be searched and retrieved. The only way to locate a record or image file without adequate description is to browse the entire collection. Descriptions must be accurate and consistent, and the vocabulary of terms used understood by both the cataloger and the user. Consistent descriptions and cataloging make it possible to link databases and search across collections to reach a wider audience.

To provide this consistency of description, libraries, archives, and museums use cataloging conventions, and thesauri of common descriptors. For searching online, catalog records with fields of data, called *metadata*, are required. One important metadata standard that has emerged is the Dublin Core, which outlines several key fields of information. Dublin Core is a simple, international standard. Please refer to the *Metadata Guidelines – September 2005* (found on the Arizona State Library, Archives and Public Records website) for specific Dublin Core-based metadata standards for digital projects. Use the *Metadata Guidelines – September 2005* if your organization is collaborating on a digital project with the Arizona State Library, Archives and Public Records.

## USING THE DUBLIN CORE FORMAT

The combination of digital information and technology in the coming years provides amazing new opportunities for Arizona's cultural and historical institutions. Every citizen of the state, as well as many beyond, will increasingly be able to view the unique artifacts, documents, writings, photographs and maps held by the various museums, archives and libraries in Arizona and throughout the world.

In order to coordinate access to the collections in many institutions, these online resources will need to be created (scanned or digitized), indexed, organized and maintained in systems which are able to work together. The goal is to provide the same access to anyone, wherever they may be – Tuba City, Phoenix or Sierra Vista. Special vocabulary, called nomenclature, exists to help create metadata. It builds on a shared structure and vocabulary guide describing library, archives or museum items or collections.

You will need to organize the information about each item that you digitize into a database to provide access and permit searching. Representatives from libraries, archives, and museums and many other professions have created a brief list of common cataloging elements to permit sharing across collections. This list of cataloging fields is called the Dublin Core. At its simplest, it is a system that creates the indexes that work together. Almost every museum artifact, archival document, photograph, or publication can fit into some of these categories. The system has been designed for international use of a broad range of disciplines, such as government, the sciences, social sciences, humanities, and business.

Many Dublin Core fields are repeatable, if necessary, to adequately describe your collection. For example: you may choose to list more than one subject heading.

We recommend that you refer to the *Metadata Guidelines – September 2005* by the Arizona State Library, Archives and Public Records for more specific information about the creation, use and application of metadata.

If your catalog already exists in a structured database which follows established standards, it may be possible to “translate” your information to the Dublin Core to eliminate reentry of data. This process is called creating a “crosswalk” from your catalog field to the Dublin Core field for searching and retrieval. Crosswalk formats are currently being developed nationally, and may not be immediately available for you to implement. A sample crosswalk for the Arizona State Library, Archives and Public Records can be found in the *Metadata Guidelines – September 2005* on our website (<http://www.lib.az.us>).

Take your time. Be thorough when describing and creating the items. The more description you give (for instance, all of the people in a photograph), the more valuable your information will be. Think creatively. A patron once asked us for pictures of hats—we had never cataloged “hats” in our database, but we watch for them now. Unusual details, like mesquite fences and peculiar shoes, as well as George Hunt's spittoon in a photograph, can be a treasure to a particular researcher. On the other hand it is very time consuming and unwise to minutely

over-describe an object or collection. Balance and good judgment are essential skills in the art of cataloging.

Above all, have fun! Looking closely at the treasures in your collection will not only provide better databases, but it will also make the experience more enjoyable.

## **USING A DATABASE**

Most organizations use a database to store cataloging information which then permits users to search and retrieve information. Databases organize the catalog information into a structured set of fields. In addition to the Dublin Core categories, it is likely that your database will include administrative information, such as donor records, which are important to track internally but may not be available to other institutions or the general public. As with cataloging, it is important that the type of database, structure and content of the fields are consistent across projects.

In the simplest terms, a database is an index. It is a tool, such as a catalog, which enables researchers to find what they seek.

The magic of a database is that it can allow for sorting of information in many ways. If your collection includes photographs, brass items, and campaign buttons, it could be sorted to find all the items you have about Barry Goldwater's presidential campaign, or all the photographs you have of Paul Fannin (even if he is included on a photograph featuring somebody else), or all the items created in 1964.

Databases can be elaborate. Libraries often have created complex databases, called catalogs, which adhere to rigid standards and MARC format. The people who catalog this information have years of training and create sophisticated records that help library users. This also allows many libraries to "pool" their resources and to share their catalogs online. The computers that drive these systems, and the software which organizes them, can frequently be extremely expensive.

Unless you are associated with a very large city, state, university or the federal government, you may need to find a way to do something similar with much less money and staff. Before you decide to purchase a database software package for your digital collection, consider contacting the Arizona State Library, Archives and Public Records to learn more about the database software used by the Arizona Memory Project.

## **SELECTING A DATABASE**

There are two approaches to database systems. The first incorporates an existing library catalog with specialty software designed for use with non-print items or special collections. These can be well integrated into the main catalog system, but should still allow the necessary flexibility for managing your collection.

Another good process is to use a good general database. Database software such as Microsoft Access or Claris Filemaker Pro can be purchased at computer stores or online and cost less



than \$500. Filemaker will operate on both Macintosh and IBM-compatible systems, while Access currently works only on PC's (IBM compatible Personal Computers).

Less expensive databases, particularly older ones, may be proprietary, meaning they are written in a special language not all software can read. Proprietary databases make the transfer of data to new versions or other software difficult or impossible.

Exportability is one of the most important aspects in any database. This means that your data can be transferred to another database, either because you upgrade to a newer type or because you are linking (connecting) your information with other institutions, as many libraries do. The time is near when libraries, archives, museums and historical societies will all pool our information to help researchers find our resources.

To ensure that your database is exportable and will work with others, look for the following:

- It can be exported in tab or comma delimited export. These are universal file types recognized by all modern databases and should continue to be recognizable in the foreseeable future.
- It supports open database connectivity (ODBC).
- Your database should be flexible, yet simple. This feature is difficult to find. You should be able to create or adapt a “template” (what you see on your screen) without being a computer programmer.
- You should be able to “define” the fields, their size, and have as many as you need. This means that you can create the database that helps you lay out Dublin Core fields, plus your own administrative information. Most databases limit the size and number of fields that can be exported in DBF and these limitations may be hidden in small print.
- It should have unlimited capabilities. Many older databases “lockout” at a designated number of entries—5,000, 10,000 or more. You may not think you will ever reach that number, but you might be surprised. When records are merged someday, you may have millions of access points on your database.
- It should be relational. You should be able to create several different databases (for example, one for the items in your collection and one for circulation) which can connect, without having to enter the information twice.

We recommend that you enroll in a workshop or consult with our Agency staff about the layout of your database. Every institution's needs are different. The basic fields explained in Dublin Core are the most important for future connections, and they probably will also be the most common ones that you will use. But you may have other needs—circulation status, provenance (the history of the item in your collection), relationships within your collection (like a spittoon that was once used by an important person—you might not normally collect spittoons, but this one belonged to Gov. Hunt, so you keep it).

Once your computer hardware and database software are purchased it is important that you devote some time upfront becoming familiar with them. There is usually a tutorial. Going through the software carefully in the beginning will help you understand the boxes and layout that you might prefer, and find easy to work with. You can move fields around so that the ones important to you are where you see them first.

When including ID numbers for digital files in a database, be sure to include the extension format on all PC computers. These should look something like .jpg or .tif. This will allow you to link them to HTML or other query languages.

Give considerable thought to the metadata elements and fields that are important to you. Refer to the Arizona State Library, Archives and Public Records *Metadata Guidelines - September 2005* for more information on recommended best practices for metadata creation. It is much better to devote time laying out the database at the beginning, than it is to go back after you've created 5000 records and realize that you didn't include the date of creation, or the physical description. Otherwise, after the fact, you might have to access each item's record and add that information.

## **PRESENTING YOUR MATERIALS**

Once your materials have been digitized, cataloged and organized, they must be made available to potential users. The access you provide might include stand-alone presentation on dedicated computers in your institution, on CD-ROM's or DVDs, or to a larger community via the World Wide Web. All types of presentation require development of computer screens that will act as an interface for users interacting with the database to select, retrieve and view material on a computer screen.

Scanning an image, writing a catalog description, or designing a Web page or database form may be simple individual activities. However, designing a large-scale web site that may contain hundreds or thousands of scans, catalog records, and layouts that work together effectively is a complicated process. Simple, effective organization and design takes expertise and time. In developing your design, remember your users and consider their experience and interests.

If this is your first project, you are encouraged to start with a small project to acquire some experience. Consider developing one or more pilot projects with only a few images to test and evaluate your design and organization. Our Agency hopes to learn from projects and to coordinate and share guidelines, tips, and "best practices" to continue to improve the scope and quality of digital collections within our state.

### **Design for Usability**

One of the first design tasks is to consider the information you will present.

Information is organized by intellectual content such as function, subject, chronology, or geography. The strategy for organizing your material needs to be consistent throughout the project to provide a logical structure for patrons to search for and identify desired information.

Screen and interface design are particularly important for the success of digital projects, especially if the audience is not familiar with computers or the Web.

As you consider the usability of your project, think about what you like and dislike about other similar projects. What made the design easy to use? What is frustrating about those designs? What is missing and should be added?

Jakob Nielsen, a national authority on the subject, has identified ten basic principles of usability:

Use simple and natural dialogue	Speak the users' language
Minimize user memory load	Be consistent
Provide feedback	Clearly mark exits
Provide shortcuts	Provide good error messages
Prevent errors	Provide help and documentation

In general, interfaces should be simple. Design should be attractive, but aesthetics should not drive interface decisions. Rather, you should ensure that your design is clear and easy to use.

Many Web designers, especially volunteers without a professional background in interface design, want to use advanced browser features. However, remember that your users may not have the most current equipment or software and may be using slow modems to access your information. Also, be aware that the way a site looks can vary across browsers, and between versions of browsers. Refer to the Web site: <http://www.w3schools.com/browsers/default.asp> for information on different browsers currently being used.

One important factor of Web usability is called navigation. Within a document, users should be able to move back to the "home" page quickly from any screen. This gives them the ability to start over quickly if they get lost or confused. You may want to provide links to both the project home as well as other intermediate pages. For example, if a user moves within a document that takes up more than one screen, you may want to provide links that enable users to go to the initial welcome screen, as well as the home page or navigation screen for the document they are currently viewing.

A multi-screen document should also permit users to simply move forward or backward. If a document requires readers to scroll off the screen, you may want to include forward and backward links at the top and bottom of the page so that users don't have to scroll to the top to get to the link for the next page.

It is important to consider the needs of specialized audiences, such as those with visual disabilities. The Americans with Disabilities Act requires reasonable accommodation of people with disabilities. When designing for the Web, you should remember that these users often rely on automatic screen readers. Options for accessibility include offering a text-only version,

limiting information presented only as graphic images, and ensuring that all graphics have an automatic text description of the graphic. A special Web site, WebXACT (<http://webxact.watchfire.com/>), will check a web address and report accessibility problems.

Many Web sites give information about usability for disabled communities. While this information is targeted towards Web design, much of it will also apply to stand-alone projects.

## **LONG-TERM STORAGE AND MAINTENANCE OF THE DIGITAL FILE**

Digital image files are not “permanent.” In fact, at present, they are among the most unstable, fragile storage media available. *Digital files should always be considered transient and will require ongoing care to protect them from obsolescence and corruption.*

There are many definitions for the word “permanent.” In the electronic world “permanent” or “archival” files may denote only a few years of longevity. Cultural institutions need to interpret this word as meaning for the very long-term - decades, centuries, indefinitely. Unfortunately, digital files are not permanent in any sense of the word, and are extremely vulnerable to deterioration and loss.

In purchasing digital systems you should be extremely wary of vendor claims of “permanence” and ease of long-term sustainability. Any extravagant claims in this area probably will not reflect current research of these complex issues, nor the rudimentary methods that are available currently. In these situations, vendors’ interests in selling you their product may not coincide with your institution’s needs. You should ask probing questions of all vendors.

For now, efforts to maintain sensitive digital information must rely on managing the integrity of the data, not the storage media. Effective project planning should include resources to migrate the data, meaning that every few years the digital files are transferred to new media, from CD to CD (refreshing) or CD to hard drive. Even this method presents hazards in file degradation.

Given the current environment, image files scanned today will eventually need to be rescanned again in the future.

Regardless of the maintenance role identified in your planning, all projects should include a management plan to back-up data files to provide for disaster recovery if the media fails or a computer crashes. Effective planning should also include resources for storage of back-up files off-site, to periodically check the integrity of the master digital files and back-up files, and to plan for migration to new media.

Technology is quickly evolving. What we regard as a “high resolution” 1200 dpi file today, may be considered unsatisfactory in a few years. Imaging software will improve, as will storage capabilities and retrieval and presentation options. Given this state of constant flux, the best method to follow is to balance the resources available and use these to determine today’s

guidelines. Planning decisions will involve your current capabilities for staff time, image resolution, storage and retrieval. You should realize that today's guidelines are not permanent, unchangable "standards." Most certainly these guidelines will evolve as we collectively learn from our projects and as the scanning and digitizing field continues to develop.

## APPENDIX A: FREQUENTLY ASKED QUESTIONS

1. Q. I work at a small institution. I want to begin a digital project but I do not have much knowledge, skill, or technical ability in this area. Would it be appropriate for me to contract this work out to a computer business that specializes in this work? If so, what information should I need to know to select and successfully work with a company?

*Yes, several institutions have opted for this method. These services are offered by larger cultural institutions and by private companies. If you choose to follow this method, it is very important that you still acquire a good background of digitization processes and that you keep up-to-date with new developments. This knowledge is necessary for making wise planning decisions and for quality control of the work performed. We recommend that you become familiar with the Web sites listed in this report. A successful project requires constant, informed communication between all participants. The most qualified contractors you choose may not be located in your area or even in Arizona. You will need to decide if you require that the work be done on site at your institution, or if materials can be temporarily moved to your contractor's work site.*

2. Q. You mentioned that in most cases digitization should not be considered a substitute for conventional archival photographic preservation methods. What are some exceptions to this?

*A. Color photographs using Ektachrome, Ansco and some other early types of color film during the 1940-1982 time period are among the most fragile photographic items in most collections. They are almost all experiencing dye instability that cannot be reversed. These photographs can be stabilized, but in many cases the deterioration is already significant. Scanning these photographs or negatives, repairing the image in photo software, and printing using the best possible archival systems (printer and paper) offers the most affordable current means of preservation. Ideally, the prints would each be photographed on more stable current print paper. Realistically, most institutions cannot afford to do this. Scanning in very high resolution files can be considered an archival option for these images and may be the only instance where the digital file is more stable than the deteriorating original.*

*Some institutions with high budgets are scanning all images with as high a ratio as possible, and consider these master files as backup preservation for their collections if calamity should befall the originals. However, this approach requires considerable funds for memory storage when considering thousands or tens of thousands of scans. Also, these institutions must be extremely concerned with the complicated issues of long-term storage and migration of these files in the future.*

3. Q. There are some oversize maps and photographs in my collections which I would like to scan. What special skills do I need to do this?

*A. Oversize, three dimensional and non-conforming items present unique challenges. While they technically can be manipulated to fit within the standards and guidelines, in reality these may be better treated separately. It may be preferable, for example, to first photograph a museum artifact, and then scan the photograph. Oversize items, particularly maps and panoramic photographs, may require multiple scans and "sewing" in the software, which can be extremely time-consuming and often shows*

*in the final file. At the very least, institutions should consider the extra time demands if their collections include large numbers of these items.*

4. Q. Do you recommend using digital cameras in the scanning process?

*A. Yes, digital cameras may be useful for organizations that have the technical and financial resources. Generally, small local institutions cannot afford them. However, their capabilities and prices are rapidly improving and may become the ideal scanning device in the future. This is very important to institutions with bound volumes of oversize documents such as newspapers, three dimensional items, and special items. Unbinding, size reduction by cutting and anything which results in the defacing of these resources in order to facilitate either microfilming or scanning at this time should be given careful consideration.*

5. Q. I understand that there are a variety of “guidelines” in use. Could I collaborate directly with another library, archives or museum that is using different guidelines other than the ones you suggest?

*A. Your question is important. For a variety of reasons, guidelines are emerging which may or may not be entirely consistent with others. There are many good reasons to cooperate with other institutions: you could possibly share equipment, staff expertise, and beginning frustrations and challenges are not as daunting if shared. Many institutions throughout the U.S. are participating in shared projects. Each of the libraries of Arizona’s three public universities, Northern Arizona University, Arizona State University, and the University of Arizona have established digitization programs. You may wish to develop a cooperative relationship with one of these libraries, or the larger public libraries or museums in the state. If so, it is important that you make contact with one of these institutions prior to beginning your project so that you will not create confusion and waste your valuable resources down the road. Refer to the planning guidelines (found on page 5) for more information.*

## APPENDIX B: RECOMMENDED RESOURCES

Many of the sources below are available at Web sites. If you do not have an Internet connection you may be able to gain access at your local public library, or contact Wendi Goen for acquiring selected paper copies.

### Dublin Core

Dublin Core Metadata Initiative

<http://purl.oclc.org/dc/>

### Cataloging and Metadata

*Metadata Guidelines – September 2005* (Arizona State Library, Archives and Public Records)

<http://www.lib.az.us>

Collaborative Digitization Program: Digital Toolbox

<http://www.cdphheritage.org/digital/index.cfm>

EAD (Encoded Archival Description) Format

<http://lcweb.loc.gov/ead/>

Gill, Tony, et. al. *Introduction to Metadata: Pathways to Digital Information* (Getty Trust Publications, 1999.)

Also available in paper from Getty Trust Publications 800-223-3431 ISBN 0-89236-533-1

[www.getty.edu/bookstore/index.html](http://www.getty.edu/bookstore/index.html)

[www.getty.edu/research/institute/standards/intrometadata/index.html](http://www.getty.edu/research/institute/standards/intrometadata/index.html)

Medeiros, N. (1999). Making room for MARC in a Dublin Core world. *Online*, 23(6), 57-60.

Milstead, Jessica and Susan Feldman. *Metadata: Cataloging by Any Other Name . . .*, *ONLINE*, January, 1999.

<http://www.cbuc.es/5digital/1.pdf>

USMARC Format

<http://lcweb.loc.gov/marc/>

### Examples of Digital Projects

American Memory Digital Project

<http://memory.loc.gov/ammem/>

Arizona Memory Project

<http://azmemory.lib.az.us>

Australian National Library

[http://amol.org.au/collection/collections\\_index.asp](http://amol.org.au/collection/collections_index.asp)

Collaborative Digitization Program

<http://www.cdphheritage.org>

Online Archive of California

<http://sunsite2.berkeley.edu/oac/>



Western Waters Digital Library  
<http://www.westernwaters.org/>

## **Crosswalking**

Dublin Core/MARC/GILS Crosswalk  
<http://www.loc.gov/marc/dccross.html>

Getty Standards Program Crosswalk of Metadata Standards  
[http://www.getty.edu/research/conducting\\_research/standards/intrometadata/3\\_crosswalks/index.html](http://www.getty.edu/research/conducting_research/standards/intrometadata/3_crosswalks/index.html)

## **Digital Project Guidelines**

A Framework of Guidance for Building Good Digital Collections  
<http://www.niso.org/framework/Framework2.html>

Collaborative Digitization Program: Digital Toolbox  
<http://www.cdphheritage.org/digital/index.cfm>

Planning Digital Projects for Historical Collections in New York State  
<http://digital.nypl.org/brochure>

National Archives and Records Administration Digital Guidelines  
<http://www.archives.gov/preservation>

The NINCH Guide to Good Practice in the Digital Representation & Management of Cultural Heritage Materials  
<http://www.ninch.org/>

## **Digitizing Photographs and Historical Documents**

Franziska S. Frey and James M. Reilly, *Digital Imaging for Photographic Collections: Foundations for Technical Standards* (Rochester, NY: Image Permanence Institute, 1999).

Townsend, Sean, et. al., *Digitising History: A Guide to Creating Digital Resources from Historical Documents* (United Kingdom: Oxbow Books, 1999). Available from the Society of American Archivists, product code 381, 312/922-0140, Ext. 21 or [publications@archivists.org](mailto:publications@archivists.org).

## **Recommended Thesaurus and Cataloging Tools For Database Building**

*Art and Architecture Thesaurus* (AAT)  
<http://www.getty.edu/research/tools/vocabulary/aat>

Getty Thesaurus of Geographic Names (TGN)  
<http://www.getty.edu/research/tools/vocabulary/tgn>

Library of Congress Thesaurus For Graphic Materials (TGM)  
<http://www.loc.gov/rr/print/tgm1/>

Library of Congress Subject Headings, 29th Edition (2006)  
(\$225 four volume paper set) Contact Library of Congress Catalog Distribution Service  
<http://catalog.loc.gov/>

Anglo-American Cataloging Rules, 2nd ed., rev. 2005 (AACR2)  
Print and online versions can be ordered at: <http://www.aacr2.org/>

## **Museum Applications and Projects**

Museums and the Online Archive of California  
<http://www.oac.cdlib.org>

Consortium for the Computer Interchange of Museum Information (CIMI)  
Guide to Best Practices: Dublin Core, August, 1999  
<http://www.cimi.org/projects/html> (not updated recently)

RLG REACH Project  
<http://www.rlg.org/>

Visual Resources Association Core Categories  
<http://www.vraweb.org>

Categories for the Description of Works of Art  
<http://www.getty.edu/research/institute/standards/cdwa/index.html>

AMOL (Australian Museums Online)  
<http://amol.org.au>

## **Developments in the Digital Resource Movement**

State of Kansas Information Technology Advisory Board – Electronic Records Committee  
<http://da.state.ks.us/itab/erc>

Links to hundreds of sources on a variety of topics related to electronic records. Supported by the NHPRC (National Historical Records and Publications Commission).

Berkeley Digital Library Sunsite. Imaging Information. Many sources and citations  
<http://sunsite.berkeley.edu/imaging/>

Background Papers and Technical Information. American Memory Project, Library of Congress.  
<http://memory.loc.gov/ammem/ftpfiles.html>

Zorich, Diane M. *Introduction to Managing Digital Assets: Options for Cultural and Educational Organizations* (Getty Information Institute, 1999).

DeWitt, Donald L., ed. *Going Digital: Strategies for Access, Preservation, and Conversion of Collections to a Digital Format* (Haworth Press, 1998). Available from the Society of American Archivists, 312/922-0140, ext. 21 or [publications@archivists.org](mailto:publications@archivists.org). Product code 354. \$40.00

D-Lib Magazine-one of the best resources for contemporary up-to-date information  
<http://www.dlib.org/>

## Information Organization

Burke, Mary. *Organization of Multimedia Resources: Principles and Practice of Information Retrieval* (Gower Publishing, 1999).

Davenport, Thomas. *Information Ecology* (Cambridge: Oxford University Press, 1997).

Rosenfeld, Louis, and Peter Morville. *Information Architecture for the World Wide Web* (Sebastopol, Calif.: O'Reilly, 1998).

## Usability and Design

Nielsen, Jakob. *Usability Engineering* (Boston: Academic Press, 1993).

Nielsen, Jakob. *UseIt.Com* (<http://www.useit.com>).

Nielsen, Jakob. *Designing Web Usability: The Practice of Simplicity* (Indianapolis: New Riders, 1999).

Nielsen's writings are interesting, readable, and often funny. His ideas are based on scientific observation of user behavior rather than opinion or divination.

Yee, Martha M. and Sharon Shatford Layne. *Improving Online Public Access Catalogs* (Chicago: American Library Association, 1998).

## Copyright

American Libraries Association – Copyright

<http://www.ala.org/ala/washoff/WOissues/copyrightb/copyright.htm>

Stanford Libraries Copyright and Fair Use

<http://fairuse.stanford.edu/>

United States Copyright Office

<http://www.copyright.gov/>